Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models

Author(s): José C. Pinheiro and Edward C. Chao

Source: *Journal of Computational and Graphical Statistics*, Mar., 2006, Vol. 15, No. 1 (Mar., 2006), pp. 58-81

Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association, Institute of Mathematical Statistics, and Interface Foundation of America

Stable URL: https://www.jstor.org/stable/27594165



# Efficient Laplacian and Adaptive Gaussian Quadrature Algorithms for Multilevel Generalized Linear Mixed Models

José C. PINHEIRO and Edward C. CHAO

Mixed-effects models have become a popular approach for the analysis of grouped data that arise in many areas as diverse as clinical trials, epidemiology, and sociology. Examples of grouped data include longitudinal data, repeated measures, and multilevel data. In the case of linear mixed-effects (LME) models, the likelihood function can be expressed in closed form, with efficient computational algorithms having been proposed for maximum likelihood and restricted maximum likelihood estimation. For nonlinear mixed-effects (NLME) models and generalized linear mixed models (GLMMs), however, the likelihood function does not have a closed form. Different likelihood approximations, with varying degrees of accuracy and computational complexity, have been proposed for these models. This article describes algorithms for one such approximation, the adaptive Gaussian quadrature (AGQ), for GLMMs which scale up efficiently to multilevel models with arbitrary number of levels. The proposed algorithms greatly reduce the computational complexity and the memory usage for approximating the multilevel GLMM likelihood, when compared to a direct application of a single-level AGQ approximation algorithm to the multilevel case. The accuracy of the associated estimates is evaluated and compared to that of estimates obtained from other approximations via simulation studies.

**Key Words:** Binomial; Grouped data; Penalized quasi-likelihood; Poisson.

## 1. INTRODUCTION

Mixed-effects models have become a popular tool for the analysis of grouped data that arise in many areas as diverse as clinical trials, epidemiology, and sociology. Examples of grouped data include longitudinal data, repeated measures, and multilevel data (in which multiple, nested classification factors are present). Linear mixed-effects (LME) (Laird and Ware 1982) and nonlinear mixed-effects (NLME) models (Pinheiro and Bates 2000) are typically used to describe grouped data for which, conditional on the random effects, the

José C. Pinheiro is Senior Associate Director of Biostatistics, Novartis Pharmaceuticals, East Hanover, NO 07936 (E-mail: jose.pinheiro@novartis.com). Edward C. Chao is Research Scientist, Insightful Corporation, 1700 Westlake Avenue North, Suite 500, Seattle, WA 98109 (E-mail: echao@insightful.com).



*Journal of Computational and Graphical Statistics, Volume 15, Number 1, Pages 58–81*
*DOI: 10.1198/106186006X96962*

response variable is assumed to follow a normal distribution. In applications where the assumption of conditional normality for the response is not reasonable (e.g., binary and count data), generalized linear mixed models (GLMMs) (McCulloch and Searle 2001) are often used instead.

Multilevel GLMMs are applied to many community sociological and educational studies, such as multistage opinion preference surveys that have repeated measures from families nested in geographical regions. Hox (2002) and Raudenbush and Bryk (2002) described and discussed several examples of application of GLMMs in these areas. Typically, penalized quasi-likelihood (PQL) and marginal quasi-likelihood (MQL) methods (Breslow and Clayton 1993) are used to analyze these studies. Such methods are also commonly used in applications of GLMMs to health and biomedical studies. Leyland and Goldstein (2001) described several examples of GLMMs for multilevel binary and count data, such as a study of testicular cancer mortality throughout Europe in the 1970s. A major concern associated with PQL and MQL is the potential bias in the estimation of model parameters in the presence of (relatively) large variance components. Rodriguez and Goldman (1995) investigated this problem using multilevel clustered data from a national survey in Guatemala with a binary outcome on the usage of prenatal care for each newborn child among women within communities in the country. Their simulated datasets, produced in accordance to the design of this survey, are used in Section 4 to investigate and compare the bias of different GLMM estimation methods.

In the case of LME models, the likelihood function can be expressed in closed form. Efficient computational algorithms have been proposed for maximum likelihood (ML) and restricted maximum likelihood (REML) estimation with these models (Lindstrom and Bates 1988; Pinheiro and Bates 2000), being the two most commonly available estimation methods in software for LME models (Littell, Milliken, Stroup, and Wolfinger 1996; Pinheiro and Bates 2000). The greater generality of NLME models and GLMMs comes at a price: their likelihood functions cannot be derived in closed form, being expressed as integrals with respect to the random effects. This makes exact likelihood estimation for these models an intractable numerical problem and, as a result, the literature contains different approximate methods, with varying degrees of accuracy and computational complexity. In this article, we concentrate on GLMMs, describing computationally efficient algorithms for approximate likelihood estimation with these models.

Initial estimation methods proposed for GLMMs, penalized quasi-likelihood (PQL), and marginal quasi-likelihood (MQL) (Breslow and Clayton 1993), are based on linearization ideas similar to the ones proposed by Lindstrom and Bates (1990) in the context of NLME models. These methods have the advantage of being simple to implement with existing software for LME models, including the case of multilevel GLMMs. However, both PQL and MQL have been reported to produce biased estimates in certain cases of practical interest, most noticeably that of binary data with few observations per group (Rodriguez and Goldman 1995). Another problem with the use of PQL and MQL in practice is that the underlying objective function, which is optimized to produce the parameter estimates, is not truly an approximation to the likelihood function. This, in particular, does not allow the

use of likelihood ratio tests to compare nested GLMM models based on the PQL or MQL objective functions.

More accurate, and computationally intensive, approximations to the NLME and GLMM likelihood have been proposed in the literature, including Laplacian and quadrature methods. First-order Laplacian approximations (Tierney and Kadane 1986; Pinheiro and Bates 1995) have been reported to produce biased estimates under some distributional scenarios, with higher order Laplacian approximations (Raudenbush, Yang, and Yosef 2000) producing more accurate estimates. Adaptive Gaussian quadrature (AGQ) methods (Pinheiro and Bates 1995) and spherical radial quadrature methods (Clarkson and Zhan 2002) can be implemented with arbitrary degrees of accuracy, leading to nearly unbiased estimates, but at the cost of increased computational complexity. All of these more accurate approximations, however, have been proposed and illustrated in the literature only for single-level GLMMs, with existing algorithms not scaling up efficiently to the multilevel case.

This article describes computationally and memory-efficient AGQ approximation algorithms for GLMMs which scale up efficiently to multilevel models with arbitrary number of levels. These algorithms rely on methods developed by Pinheiro and Bates (2000) for highly efficient computation of the multilevel LME likelihood, adapting and extending them to the GLMM context. The resulting algorithms greatly reduce the computational complexity of approximating the multilevel GLMM likelihood, when compared to a direct application of a single-level AGQ approximation algorithm to the multilevel case. Noting that the first-order Laplacian approximation is just a one-point AGQ approximation, the proposed algorithms also apply to this approximation.

The rest of the article is organized as follows. Section 2 describes the Laplacian and AGQ approximation algorithms for single-level GLMMs. Section 3 extends these algorithms for multilevel GLMMs. Section 4 evaluates the accuracy of the Laplacian and AGQ approximations and compares it to that of PQL, using simulated datasets provided by Rodriguez and Goldman (1995), as well as our own simulated data. Section 5 gives conclusions and further considerations.

## 2. SINGLE-LEVEL GENERALIZED LINEAR MIXED MODELS

This section considers GLMMs for grouped data with a single classification factor. We present a brief review of the model, introducing the notation that will be used throughout the article, and discuss likelihood estimation in the context of these models. The Laplacian and adaptive Gaussian quadrature (AGQ) approximations to the marginal log-likelihood function of the single-level GLMM model are described. Applications of the algorithms to the binomial and Poisson families are used for illustration.

### 2.1 MODEL DEFINITION AND NOTATION

The basic formulation of a single-level GLMM is that the response vector $\mathbf{y}_i$ for a given group $i$, conditional on random effects $\mathbf{b}_i$, is distributed as independent random variables with distribution in the exponential family. The most important cases of practical interest

are the binomial and Poisson distributions.

Given $\mathbf{b}_i$, the conditional density of $\mathbf{y}_i$ in canonical form is

$$\begin{aligned} p(\mathbf{y}_i|\mathbf{b}_i) &= \prod_{j=1}^{n_i} \exp\left\{\left(y_{ij}\theta_{ij} - d(\theta_{ij})\right)/a(\phi) + c(y_{ij},\phi)\right\} \\ &= \exp\left\{\left(\mathbf{y}_i'\boldsymbol{\theta}_i - d(\boldsymbol{\theta}_i)'\mathbf{1}\right)/a(\phi) + c(\mathbf{y}_i,\phi)'\mathbf{1}\right\}, \end{aligned}$$

for appropriate functions $a(.)$, $d(.)$, and $c(.)$. The random effects are assumed to be independently distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$.

The model is further determined through the specification of an invertible *link* function $g(.)$ relating the expected value of $\mathbf{y}_i$, $\boldsymbol{\mu}_i = \mathrm{E}(\mathbf{y}_i|\mathbf{b}_i)$, to a set of covariates and the fixed and random effects. That is,

$$g(\boldsymbol{\mu}_i) = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i = \boldsymbol{\eta}_i \quad \text{and} \quad \boldsymbol{\mu}_i = h(\boldsymbol{\eta}_i) = h(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i),$$

where $\boldsymbol{\beta}$ denotes the vector of fixed effects, $\mathbf{X}_i$ and $\mathbf{Z}_i$ are the (known) fixed and random effects regression matrices, and $h = g^{-1}$ is the inverse link function. We will assume the canonical link function for the exponential family, in which case $\boldsymbol{\theta}_i = \boldsymbol{\eta}_i$. It then follows that the joint density of $(\mathbf{y}_i, \mathbf{b}_i)$ is given by

$$\begin{aligned} p(\mathbf{y}_i, \mathbf{b}_i) = \exp\Big\{ &\left[\mathbf{y}_i'\left(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i\right) - d\left(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i\right)'\mathbf{1}\right] \\ &/a(\phi) + c\left(\mathbf{y}_i,\phi\right)'\mathbf{1} - \mathbf{b}_i'\boldsymbol{\Psi}^{-1}\mathbf{b}_i/2\Big\} \Big/ \left[(2\pi)^{q/2}|\boldsymbol{\Psi}|^{1/2}\right], \end{aligned}$$

where $q$ denotes the number of random effects (the length of $\mathbf{b}_i$).

As with any mixed-effects models, because the random effects are nonobservable quantities, likelihood estimation must rely on the *marginal* density of $\mathbf{y}_i$, which is obtained by integrating the joint likelihood function with respect to $\mathbf{b}_i$. Like in NLME models, for GLMMs this integral does not have a closed form expression and approximations are required for computationally feasible estimation. The next sections describe two such methods, the Laplacian and adaptive Gaussian quadrature (AGQ) approximations, discussing efficient computational methods for their implementation.

## 2.2 LAPLACIAN APPROXIMATION

Laplacian approximations are frequently used in Bayesian inference to estimate marginal posterior densities and predictive distributions (Tierney and Kadane 1986; Leonard, Hsu, and Tsui 1989). The use of the Laplacian approximation for single-level NLME models was described by Pinheiro and Bates (1995) and for multilevel NLME models in Pinheiro and Bates (2000). These techniques can also be used for approximating the log-likelihood function in GLMMs, as described in the following. We consider in this section the single-level case, for which the (marginal) GLMM likelihood can be expressed

as

$$\begin{aligned} p(\mathbf{y}_i) &= \int p(\mathbf{y}_i, \mathbf{b}_i) d\mathbf{b}_i \\ &= (2\pi)^{-q/2} |\boldsymbol{\Psi}|^{-1/2} \int \exp\left[g(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, \mathbf{y}_i, \mathbf{b}_i)\right] d\mathbf{b}_i, \\ g(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, \mathbf{y}_i, \mathbf{b}_i) &= \left[\mathbf{y}_i'(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i) - d(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i)'\mathbf{1}\right] \\ &\quad /a(\phi) - \mathbf{b}_i'\boldsymbol{\Psi}^{-1}\mathbf{b}_i/2 + c(\mathbf{y}_i, \phi)'\mathbf{1} \\ &= \left[\mathbf{y}_i'\boldsymbol{\eta}_i - d(\boldsymbol{\eta}_i)'\mathbf{1}\right]/a(\phi) - \mathbf{b}_i'\boldsymbol{\Psi}^{-1}\mathbf{b}_i/2 + c(\mathbf{y}_i, \phi)'\mathbf{1}. \end{aligned}$$

The idea behind the Laplacian and, to some extent, also the adaptive Gaussian approximation, is to approximate $g(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, \mathbf{y}_i, \mathbf{b}_i)$ by a second-order Taylor expansion around the value of $\mathbf{b}_i$ that maximizes $g(., \mathbf{b}_i)$. Note that

$$\begin{aligned} \frac{\partial g(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, \mathbf{y}_i, \mathbf{b}_i)}{\partial \mathbf{b}_i} &= \mathbf{Z}_i^T \left[\mathbf{y}_i - d'(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i)\right]/a(\phi) - \boldsymbol{\Psi}^{-1}\mathbf{b}_i \\ &= \mathbf{Z}_i^T \left(\mathbf{y}_i - \boldsymbol{\mu}_i\right)/a(\phi) - \boldsymbol{\Psi}^{-1}\mathbf{b}_i \qquad (2.1) \\ \frac{\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, \mathbf{y}_i, \mathbf{b}_i)}{\partial \mathbf{b}_i \partial \mathbf{b}_i^T} &= -\left(\mathbf{Z}_i^T d''(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i)\mathbf{Z}_i/a(\phi) + \boldsymbol{\Psi}^{-1}\right) \\ &= -\left(\mathbf{Z}_i^T \frac{\mathbf{V}_i}{a^2(\phi)}\mathbf{Z}_i + \boldsymbol{\Psi}^{-1}\right), \qquad (2.2) \end{aligned}$$

where $\mathbf{V}_i = \text{var}(\mathbf{y}_i|\mathbf{b}_i)$ represents the diagonal matrix with elements $\text{var}(y_{ij}|\mathbf{b}_i)$, and $d'$, $d''$ represent respectively the first- and second-order derivatives. It follows from (2.2) that $\partial^2 g(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi\mathbf{y}_i, \mathbf{b}_i)/\partial \mathbf{b}_i \partial \mathbf{b}_i'$ is negative-definite and, as a result, $g(., \mathbf{b}_i)$ is a strictly concave function of $\mathbf{b}_i$. Therefore, there is a unique point of maximum $\widehat{\mathbf{b}}_i$ corresponding to $\partial\, g(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, \mathbf{y}_i, \mathbf{b}_i)/\partial \mathbf{b}_i|_{\mathbf{b}_i = \widehat{\mathbf{b}}_i} = \mathbf{0}$. Equation (2.1) provides a simple recurrence relation for determining $\widehat{\mathbf{b}}_i$

$$\widehat{\mathbf{b}}_i^{k+1} = \boldsymbol{\Psi}\mathbf{Z}_i' \left(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}_i^k\right)/a(\phi), \quad \widehat{\boldsymbol{\mu}}_i^k = h\left(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\widehat{\mathbf{b}}_i^k\right),$$

where the iterations start at some $\widehat{\mathbf{b}}_i^0$, which, in the absence of prior estimates, can be set to $\mathbf{0}$. Quicker convergence is achieved using a Newton-Raphson algorithm, using the gradient vector and the matrix of second derivatives

$$\widehat{\mathbf{b}}_i^{k+1} = \widehat{\mathbf{b}}_i^k + \left(\mathbf{Z}_i' \frac{\mathbf{V}_i}{a^2(\phi)}\mathbf{Z}_i + \boldsymbol{\Psi}^{-1}\right)^{-1} \left(\mathbf{Z}_i' \frac{(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}_i^k)}{a(\phi)} - \boldsymbol{\Psi}^{-1}\widehat{\mathbf{b}}_i^k\right). \qquad (2.3)$$

A more computationally efficient version of (2.3) can be obtained by translating it into a least-squares (LS) problem. Let the augmented $\mathbf{Z}_i^{\text{aug}}$ matrix and working residuals $\mathbf{r}_i$ be defined as

$$\mathbf{Z}_i^{\text{aug}} = \begin{bmatrix} \frac{\mathbf{V}_i^{1/2}}{a(\phi)}\mathbf{Z}_i \\ \boldsymbol{\Delta} \end{bmatrix} \quad \text{and} \quad \mathbf{r}_i = \begin{bmatrix} \mathbf{V}_i^{-1/2}\left(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}_i^k\right) \\ -\boldsymbol{\Delta}\widehat{\mathbf{b}}_i^k \end{bmatrix},$$

where $\mathbf{V}_i^{1/2}$ denotes the diagonal matrix with diagonal elements given by $[\text{var}(y_{ij}|\mathbf{b}_i)]^{1/2}$ and $\boldsymbol{\Delta}$ denotes a precision matrix satisfying $\boldsymbol{\Delta}'\boldsymbol{\Delta} = \boldsymbol{\Psi}^{-1}$. It can be verified that the increment $\widehat{\delta}_i^{k+1} = \widehat{\mathbf{b}}_i^{k+1} - \widehat{\mathbf{b}}_i^k$ corresponds to the LS solution of $\mathbf{r}_i = \mathbf{Z}_i^{\text{aug}}\boldsymbol{\delta}$. Let $\mathbf{Z}_i^{\text{aug}} = \mathbf{Q}_i\mathbf{R}_i$ represent the Q-R decomposition (Thisted 1988) of $\mathbf{Z}_i^{\text{aug}}$. A computationally efficient implementation of (2.3) is then given by

$$\widehat{\mathbf{b}}_i^{k+1} = \widehat{\mathbf{b}}_i^k + \mathbf{R}_i^{-1}\mathbf{Q}_i'\mathbf{r}_i.$$

Note also that $\mathbf{Z}_i' \frac{\mathbf{V}_i}{a^2(\phi)}\mathbf{Z}_i + \boldsymbol{\Psi}^{-1} = \mathbf{R}_i'\mathbf{R}_i$.

The Laplacian approximation to the single-level GLMM log-likelihood corresponding to group $i$ is then given by

$$\ell_{\text{Lap}}(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi|\mathbf{y}_i) = \log|\boldsymbol{\Delta}| - \log|\mathbf{R}_i| + g\left(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, \mathbf{y}_i, \widehat{\mathbf{b}}_i\right). \tag{2.4}$$

The Laplacian approximation for the full data is then simply $\ell_{\text{Lap}}(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi|\mathbf{y}) = \sum_{i=1}^{M} \ell_{\text{Lap}}(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi|\mathbf{y}_i)$.

For illustration, we consider the particular cases of the binomial and Poisson distributions, for which $a(\phi) = 1$. In the case of the binomial distribution, letting $\mathbf{n}_i$ represent the vector with the number of trials corresponding to each binomial outcome in group $i$, $\mathbf{V}_i = \text{diag}\left\{\left[\mathbf{n}_i\widehat{\boldsymbol{\mu}}_i(\mathbf{1} - \widehat{\boldsymbol{\mu}}_i)\right]\right\}$, $d(\theta) = n\log[1 + \exp(\theta)]$, and $c(\mathbf{y}_i, \phi)'\mathbf{1} = \sum_j \log\binom{n_{ij}}{y_{ij}}$, so that (2.4) simplifies to

$$\begin{aligned}\ell_{\text{Lap}}^{\text{Bin}}(\boldsymbol{\beta}, \boldsymbol{\Psi}|\mathbf{y}_i) = \log|\boldsymbol{\Delta}| - \log|\mathbf{R}_i| + \mathbf{y}_i'(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\widehat{\mathbf{b}}_i) \\ -\mathbf{n}_i'\log\left[\mathbf{1} + \exp(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\widehat{\mathbf{b}}_i)\right] - \frac{1}{2}\left\|\boldsymbol{\Delta}\widehat{\mathbf{b}}_i\right\|^2 + \sum_j \log\binom{n_{ij}}{y_{ij}}.\end{aligned}$$

In the Poisson case, $\mathbf{V}_i = \text{diag}(\widehat{\boldsymbol{\mu}}_i)$, $d(\theta) = \exp(\theta)$ and $c(\mathbf{y}_i, \phi)'\mathbf{1} = -\sum_j \log(y_{ij}!)$, giving

$$\begin{aligned}\ell_{\text{Lap}}^{\text{Pois}}(\boldsymbol{\beta}, \boldsymbol{\Psi}|\mathbf{y}_i) = \log|\boldsymbol{\Delta}| - \log|\mathbf{R}_i| + \mathbf{y}_i'(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\widehat{\mathbf{b}}_i) \\ -\exp(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\widehat{\mathbf{b}}_i)'\mathbf{1} - \frac{1}{2}\left\|\boldsymbol{\Delta}\widehat{\mathbf{b}}_i\right\|^2 - \sum_j \log(y_{ij}!).\end{aligned}$$

As before, the Laplacian approximation for the full data is obtained by adding the individual Laplacian approximations for each group.

## 2.3 ADAPTIVE GAUSSIAN QUADRATURE APPROXIMATION

Gaussian quadrature rules are used to approximate integrals of functions with respect to a given kernel by a weighted average of the integrand evaluated at predetermined abscissas. The weights and abscissas used in Gaussian quadrature rules for the most common kernels can be obtained from the tables of Abramowitz and Stegun (1964) or by using an algorithm proposed by Golub (1973) (see also Golub and Welsch 1969). Gaussian quadrature rules for multiple integrals are known to be numerically complex (Davis and Rabinowitz 1984),

but, using the structure of the integrand in GLMMs, we can transform the problem into successive applications of simple one-dimensional Gaussian quadrature rules. In this section we consider the single-level GLMM.

A natural candidate for the kernel function for the quadrature rule in the single-level GLMM is the marginal distribution of the random effects, that is, the $\mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$ density. The Gaussian quadrature rule in this case can be viewed as a deterministic version of a Monte Carlo integration algorithm in which random samples of the random effects, $\mathbf{b}_i$, are generated from the $\mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$ distribution. The samples and the weights in the Gaussian quadrature rule are fixed beforehand, while in Monte Carlo integration they are left to random choice. Because importance sampling tends to be much more efficient than simple Monte Carlo integration (Geweke 1989), we consider an importance sampling version of the Gaussian quadrature rule, denoted *adaptive Gaussian quadrature* (AGQ) (Pinheiro and Bates 1995).

The critical step for the success of importance sampling is the choice of an importance distribution that approximates the integrand. For the single-level GLMM the integrand is proportional to $\exp\left[g\left(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, \mathbf{y}_i, \mathbf{b}_i\right)\right]$ which is approximated by a $\mathcal{N}(\widehat{\mathbf{b}}_i, \mathbf{R}_i^{-1}\mathbf{R}_i^{-T})$ density, with $g\left(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, \mathbf{y}_i, \mathbf{b}_i\right)$, $\mathbf{R}_i$, and $\widehat{\mathbf{b}}_i$ defined as in the previous section. This is the importance distribution used in the AGQ rule, so that the grid of abscissas in the $\mathbf{b}_i$ scale is centered around the conditional modes $\widehat{\mathbf{b}}_i$ and $\mathbf{R}_i$ is used for scaling. Let $z_j, w_j \; j = 1, \ldots, N_{\text{GQ}}$ denote, respectively, the abscissas and the weights for the (one-dimensional) Gaussian quadrature rule with $N_{\text{GQ}}$ points based on the $\mathcal{N}(0, 1)$ kernel. Define $\mathbf{z}_{\mathbf{j}} = \left(z_{j_1}, \ldots, z_{j_q}\right)^T$, $\widetilde{\mathbf{b}}_{i\mathbf{j}} = \widehat{\mathbf{b}}_i + \mathbf{R}_i^{-1}\mathbf{z}_{\mathbf{j}}$, and $W_{\mathbf{j}} = \exp(\|\mathbf{z}_{\mathbf{j}}\|^2) \prod_{k=1}^{q} w_{j_k}$. The AGQ rule is then given by

$$
\begin{aligned}
&\int \exp\left[g\left(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, \mathbf{y}_i, \mathbf{b}_i\right)\right] d\mathbf{b}_i \\
&\quad = \int |\mathbf{R}_i|^{-1} \exp\left[g\left(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, \mathbf{y}_i, \widehat{\mathbf{b}}_i + \mathbf{R}_i^{-1}\mathbf{z}\right) + \|\mathbf{z}\|^2/2\right] \exp\left(-\|\mathbf{z}\|^2/2\right) d\mathbf{z} \\
&\quad \simeq (2\pi)^{q/2} |\mathbf{R}_i|^{-1} \sum_{j_1=1}^{N_{\text{GQ}}} \cdots \sum_{j_q=1}^{N_{\text{GQ}}} \exp\left[g\left(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, \mathbf{y}_i, \widetilde{\mathbf{b}}_{i\mathbf{j}}\right)\right] W_{\mathbf{j}}.
\end{aligned}
$$

The AGQ approximation to the log-likelihood function in the single-level GLMM is then

$$
\begin{aligned}
\ell_{\text{AGQ}}\left(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi \mid \mathbf{y}\right) = M \log |\boldsymbol{\Delta}| \\
+ \sum_{i=1}^{M} \left( - \log |\mathbf{R}_i| + \log \left\{ \sum_{\mathbf{j}}^{N_{\text{GQ}}} \exp\left[g\left(\boldsymbol{\beta}, \boldsymbol{\Psi}, \phi, \mathbf{y}_i, \widetilde{\mathbf{b}}_{i\mathbf{j}}\right)\right] W_{\mathbf{j}} \right\} \right).
\end{aligned}
$$

The one point (i.e., $N_{\text{GQ}} = 1$) AGQ approximation is simply the Laplacian approximation described in the previous section, because in this case $\mathbf{z}_1 = \mathbf{0}$, $w_1 = 1$, and $W_1 = 1$.

The AGQ approximation can be made arbitrarily accurate by increasing the number of abscissas, $N_{\text{GQ}}$. However, because $N_{\text{GQ}}^q$ grid points are used to calculate the adaptive Gaussian quadrature for each group, it quickly becomes prohibitively computationally intensive as the number of abscissas increases. In practice $N_{\text{GQ}} \leq 7$ generally suffices and $N_{\text{GQ}} = 1$ often provides a reasonable approximation (Pinheiro and Bates 1995). Note

that the $(\mathbf{z}_j, W_j)$ depend only on $N_{\text{GQ}}$ and $q$, being the same for all groups. They can be calculated once and stored to be used in the calculation of the log-likelihood approximation.

In the case of the binomial distribution, the AGQ approximation simplifies to

$$\begin{aligned}\ell_{\text{AGQ}}^{\text{Bin}}\left(\boldsymbol{\beta}, \boldsymbol{\Psi} \mid \mathbf{y}\right) &= M \log|\boldsymbol{\Delta}| - \sum_{i=1}^{M}\left[\log|\mathbf{R}_i| - \sum_j \log\binom{n_{ij}}{y_{ij}}\right] \\ &\quad + \sum_{i=1}^{M} \log\left(\sum_{\mathbf{j}}^{N_{\text{GQ}}} \exp\left\{\mathbf{y}_i'\left[\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\widetilde{\mathbf{b}}_{i\mathbf{j}}\right]\right.\right. \\ &\quad \left.\left. - \mathbf{n}_i' \log\left\{\mathbf{1} + \exp\left[\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\widetilde{\mathbf{b}}_{i\mathbf{j}}\right]\right\} - \frac{1}{2}\left\|\boldsymbol{\Delta}\widetilde{\mathbf{b}}_{i\mathbf{j}}\right\|^2\right\} W_{\mathbf{j}}\right)\end{aligned}$$

and in the Poisson case

$$\ell_{\text{AGQ}}^{\text{Pois}}\left(\boldsymbol{\beta}, \boldsymbol{\Psi} \mid \mathbf{y}\right) = M \log|\boldsymbol{\Delta}| - \sum_{i=1}^{M}\left[\log|\mathbf{R}_i| + \sum_j \log(y_{ij}!)\right]$$

$$+ \sum_{i=1}^{M} \log\left(\sum_{\mathbf{j}}^{N_{\text{GQ}}} \exp\left\{\mathbf{y}_i'\left[\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\widetilde{\mathbf{b}}_{i\mathbf{j}}\right] - \exp\left[\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\widetilde{\mathbf{b}}_{i\mathbf{j}}\right]'\mathbf{1} - \frac{1}{2}\left\|\boldsymbol{\Delta}\widetilde{\mathbf{b}}_{i\mathbf{j}}\right\|^2\right\} W_{\mathbf{j}}\right).$$

## 3. MULTILEVEL GENERALIZED LINEAR MIXED MODELS

The single-level GLMM described in Section 2 can be extended to data grouped according to multiple, nested factors by modifying the models for the link functions. For example, in the case of two levels of grouping and letting $\mathbf{b}_i^{(1)}$ and $\mathbf{b}_i^{(2)} = (\mathbf{b}_{i1}', \mathbf{b}_{i2}', \ldots, \mathbf{b}_{iM_i}')'$ denote, respectively, the first and second level random effects corresponding to the Level 1 group $i$, we would have

$$\begin{aligned}p(\mathbf{y}_i|\mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)}) &= \prod_{j=1}^{M_i}\prod_{k=1}^{n_{ij}} \exp\left\{\left(y_{ijk}\theta_{ijk} - d(\theta_{ijk})\right)/a(\phi) + c(y_{ijk}, \phi)\right\} \\ &= \prod_{j=1}^{M_i} \exp\left\{\left(\mathbf{y}_{ij}'\boldsymbol{\theta}_{ij} - d(\boldsymbol{\theta}_{ij})'\mathbf{1}\right)/a(\phi) + c(\mathbf{y}_{ij}, \phi)'\mathbf{1}\right\},\end{aligned}$$

for appropriate functions $a(.)$, $d(.)$, and $c(.)$. The Level 1 random effects $\mathbf{b}_i^{(1)}$ are assumed to be independently distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_1)$ and the second level random effects $\mathbf{b}_{ij}^{(2)}$ are assumed to be independently distributed as $\mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_2)$, independent of the $\mathbf{b}_i^{(1)}$.

As in the single-level case, the model is further determined through the specification of an invertible *link* function $g(.)$ relating the expected value of $\mathbf{y}_i$, $\boldsymbol{\mu}_i = E(\mathbf{y}_i|\mathbf{b}_i)$ to a set of covariates and the fixed and random effects. That is,

$$\begin{aligned}g(\boldsymbol{\mu}_i) &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i^{(1)}\mathbf{b}_i^{(1)} + \mathbf{Z}_i^{(2)}\mathbf{b}_i^{(2)} = \boldsymbol{\eta}_i; \\ \boldsymbol{\mu}_i &= h(\boldsymbol{\eta}_i) = h\left(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i^{(1)}\mathbf{b}_i^{(1)} + \mathbf{Z}_i^{(2)}\mathbf{b}_i^{(2)}\right),\end{aligned} \tag{3.1}$$

where, as before, $\boldsymbol{\beta}$ denotes the fixed effects, $\mathbf{X}_i$, $\mathbf{Z}_i^{(1)}$, and $\mathbf{Z}_i^{(2)}$ are the (known) fixed and random effects regression matrices and $h = g^{-1}$ is the inverse link function. We will assume the canonical link function for the exponential family, in which case $\boldsymbol{\theta}_i = \boldsymbol{\eta}_i$. The Level 2 random effects matrix $\mathbf{Z}_i^{(2)}$ is block-diagonal with matrices $\mathbf{Z}_{i1}^{(2)}, \ldots, \mathbf{Z}_{iM_i}^{(2)}$ in the main diagonal. Equation (3.1) can be re-written as

$$\begin{aligned} g(\boldsymbol{\mu}_{ij}) &= \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{(1)}\mathbf{b}_i^{(1)} + \mathbf{Z}_{ij}^{(2)}\mathbf{b}_{ij}^{(2)} = \boldsymbol{\eta}_{ij}; \\ \boldsymbol{\mu}_{ij} &= h(\boldsymbol{\eta}_{ij}) = h\left(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{(1)}\mathbf{b}_i^{(1)} + \mathbf{Z}_{ij}^{(2)}\mathbf{b}_{ij}^{(2)}\right), \end{aligned}$$

where $\mathbf{y}_{ij}$, $\mathbf{X}_{ij}$, and $\mathbf{Z}_{ij}^{(1)}$ denote the subsets of $\mathbf{y}_i$, $\mathbf{X}_i$, and $\mathbf{Z}_i^{(1)}$ corresponding to the observations in $\mathbf{Z}_{ij}^{(2)}$.

Under the assumptions above, the joint density of $(\mathbf{y}_i, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)})$ is given by

$$\begin{aligned} p\left(\mathbf{y}_i, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)}\right) &= \exp\left[g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_i, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)})\right] \\ &\quad / \left[(2\pi)^{(q_1 + M_i q_2)/2} |\boldsymbol{\Psi}_1|^{1/2} |\boldsymbol{\Psi}_2|^{M_i/2}\right] \end{aligned} \tag{3.2}$$

with

$$\begin{aligned} & g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_i, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)}) \\ &= \left[\mathbf{y}_i'(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i^{(1)}\mathbf{b}_i^{(1)} + \mathbf{Z}_i^{(2)}\mathbf{b}_i^{(2)}) - d(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i^{(1)}\mathbf{b}_i^{(1)} + \mathbf{Z}_i^{(2)}\mathbf{b}_i^{(2)})'\mathbf{1}\right] \\ &\quad /a(\phi) - \mathbf{b}_i^{(1)'}\boldsymbol{\Psi}_1^{-1}\mathbf{b}_i^{(1)}/2 - \sum_{j=1}^{M_i} \mathbf{b}_{ij}^{(2)'}\boldsymbol{\Psi}_2^{-1}\mathbf{b}_{ij}^{(2)}/2 + c(\mathbf{y}_i, \phi)'\mathbf{1} \\ &= \left[\mathbf{y}_i'\boldsymbol{\eta}_i - d(\boldsymbol{\eta}_i)'\mathbf{1}\right]/a(\phi) - \mathbf{b}_i^{(1)'}\boldsymbol{\Psi}_1^{-1}\mathbf{b}_i^{(1)}/2 - \sum_{j=1}^{M_i} \mathbf{b}_{ij}^{(2)'}\boldsymbol{\Psi}_2^{-1}\mathbf{b}_{ij}^{(2)}/2 + c(\mathbf{y}_i, \phi)'\mathbf{1}, \end{aligned} \tag{3.3}$$

where $M_i$ denotes the number of Level 2 groups within Level 1 group $i$, $q_1$ denotes the number of Level 1 random effects and $q_2$ the number of Level 2 random effects.

Noting that $\mathbf{y}_i'\mathbf{Z}_i^{(2)}\mathbf{b}_i^{(2)} = \sum_{j=1}^{M_i} \mathbf{y}_{ij}'\mathbf{Z}_{ij}^{(2)}\mathbf{b}_{ij}^{(2)}$ and $d(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i^{(1)}\mathbf{b}_i^{(1)} + \mathbf{Z}_i^{(2)}\mathbf{b}_i^{(2)})'\mathbf{1} = \sum_{j=1}^{M_i} d(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{(1)}\mathbf{b}_i^{(1)} + \mathbf{Z}_{ij}^{(2)}\mathbf{b}_{ij}^{(2)})'\mathbf{1}$ and by rearranging the order of the terms in (3.3), $g_1(\cdot)$ can be factored as

$$\begin{aligned} & g_1\left(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_i, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)}\right) \\ &\qquad = g_{21}\left(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \mathbf{y}_i, \mathbf{b}_i^{(1)}\right) + \sum_{j=1}^{M_i} g_{22}\left(\boldsymbol{\beta}, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_{ij}, \mathbf{b}_i^{(1)}, \mathbf{b}_{ij}^{(2)}\right), \end{aligned}$$

where

$$g_{21}(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \mathbf{y}_i, \mathbf{b}_i^{(1)}) = \mathbf{y}_i'\left(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i^{(1)}\mathbf{b}_i^{(1)}\right)/a(\phi) - \mathbf{b}_i^{(1)'}\boldsymbol{\Psi}_1^{-1}\mathbf{b}_i^{(1)}/2 + c(\mathbf{y}_i, \phi)'\mathbf{1}$$

and

$$g_{22}(\boldsymbol{\beta}, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_{ij}, \mathbf{b}_i^{(1)}, \mathbf{b}_{ij}^{(2)}) = \left[\mathbf{y}_{ij}'\mathbf{Z}_{ij}^{(2)}\mathbf{b}_{ij}^{(2)} - d\left(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{(1)}\mathbf{b}_i^{(1)} + \mathbf{Z}_{ij}^{(2)}\mathbf{b}_{ij}^{(2)}\right)'\mathbf{1}\right] /a(\phi) - \mathbf{b}_{ij}^{(2)\prime}\boldsymbol{\Psi}_2^{-1}\mathbf{b}_{ij}^{(2)}/2.$$

It then follows that (3.2) can be re-expressed as

$$\begin{aligned} p(\mathbf{y}_i, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)}) &= \exp\left[g_{21}(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \phi, \mathbf{y}_i, \mathbf{b}_i^{(1)})\right] \Big/ \left[(2\pi)^{q_1/2}|\boldsymbol{\Psi}_1|^{1/2}\right] \\ &\quad \times \prod_{j=1}^{M_i} \exp\left[g_{22}(\boldsymbol{\beta}, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_{ij}, \mathbf{b}_i^{(1)}, \mathbf{b}_{ij}^{(2)})\right] \Big/ \left[(2\pi)^{q_2/2}|\boldsymbol{\Psi}_2|^{1/2}\right]. \end{aligned}$$

Generalizations to larger number of levels is straightforward, but the notation becomes increasingly cumbersome. For example, for three levels of nesting, letting $\mathbf{b}_i^{(1)} \overset{\text{ind}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_1)$, $\mathbf{b}_{ij}^{(2)} \overset{\text{ind}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_2)$, and $\mathbf{b}_{ijk}^{(3)} \overset{\text{ind}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_3)$ denote the first, second, and third level random effects, the joint density of $(\mathbf{y}_i, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)}, \mathbf{b}_i^{(3)})$ is

$$\begin{aligned} p(\mathbf{y}_i, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)}, \mathbf{b}_i^{(3)}) &= \exp\left[g_{31}(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \phi, \mathbf{y}_i, \mathbf{b}_i^{(1)})\right] \Big/ \left[(2\pi)^{q_1/2}|\boldsymbol{\Psi}_1|^{1/2}\right] \\ &\quad \times \prod_{j=1}^{M_i} \left\{\exp\left[g_{32}(\boldsymbol{\beta}, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_{ij}, \mathbf{b}_{ij}^{(2)})\right] \Big/ \left[(2\pi)^{q_2/2}|\boldsymbol{\Psi}_2|^{1/2}\right]\right. \\ &\quad \times \prod_{k=1}^{M_{ij}} \exp\left[g_{33}(\boldsymbol{\beta}, \boldsymbol{\Psi}_3, \phi, \mathbf{y}_{ij}, \mathbf{b}_i^{(1)}, \mathbf{b}_{ij}^{(2)}, \mathbf{b}_{ijk}^{(3)})\right] \\ &\quad \left. \Big/ \left[(2\pi)^{q_3/2}|\boldsymbol{\Psi}_3|^{1/2}\right]\right\}, \end{aligned}$$

where

$$g_{31}(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \phi, \mathbf{y}_i, \mathbf{b}_i^{(1)}) = \left(\mathbf{y}_i'(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i^{(1)}\mathbf{b}_i^{(1)})\right)/a(\phi) - \mathbf{b}_i^{(1)\prime}\boldsymbol{\Psi}_1^{-1}\mathbf{b}_i^{(1)}/2 + c(\mathbf{y}_i, \phi)'\mathbf{1},$$

$$g_{32}(\boldsymbol{\beta}, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_{ij}, \mathbf{b}_{ij}^{(2)}) = \mathbf{y}_{ij}'\mathbf{Z}_{ij}^{(2)}\mathbf{b}_{ij}^{(2)}/a(\phi) - \mathbf{b}_{ij}^{(2)\prime}\boldsymbol{\Psi}_2^{-1}\mathbf{b}_{ij}^{(2)}/2,$$

and

$$\begin{aligned} &g_{33}\left(\boldsymbol{\beta}, \boldsymbol{\Psi}_3, \phi, \mathbf{y}_{ij}, \mathbf{b}_i^{(1)}, \mathbf{b}_{ij}^{(2)}, \mathbf{b}_{ijk}^{(3)}\right) \\ &\quad = \left(\mathbf{y}_{ijk}'\mathbf{Z}_{ijk}^{(3)}\mathbf{b}_{ijk}^{(3)} - d(\mathbf{X}_{ijk}\boldsymbol{\beta} + \mathbf{Z}_{ijk}^{(1)}\mathbf{b}_i^{(1)} + \mathbf{Z}_{ijk}^{(2)}\mathbf{b}_{ij}^{(2)} + \mathbf{Z}_{ijk}^{(3)}\mathbf{b}_{ijk}^{(3)})'\mathbf{1}\right)/a(\phi) \\ &\quad\quad -\mathbf{b}_{ijk}^{(3)\prime}\boldsymbol{\Psi}_3^{-1}\mathbf{b}_{ijk}^{(3)}/2. \end{aligned}$$

We consider next the multilevel versions of the Laplacian and AGQ approximations discussed in Section 2.

### 3.1 LAPLACIAN APPROXIMATION

As before, we concentrate on the case of two-level GLMMs, with generalizations to further levels of nesting being straightforward, but notationally cumbersome. Similarly to the single-level case, we express the marginal density of $\mathbf{y}_i$ as

$$\begin{aligned} p(\mathbf{y}_i) &= \int p(\mathbf{y}_i, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)}) d\mathbf{b}_i^{(2)} \\ d\mathbf{b}_i^{(1)} &= \kappa_i \int \exp\left[ g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_i, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)}) \right] d\mathbf{b}_i^{(2)} d\mathbf{b}_i^{(1)}, \end{aligned} \tag{3.4}$$

where $\kappa_i = 1/[(2\pi)^{(q_1+M_i q_2)/2} |\boldsymbol{\Psi}_1|^{1/2} |\boldsymbol{\Psi}_2|^{M_i/2}]$. As in the single-level case, we consider second-order Taylor expansions of $g_1$ around values $\widehat{\mathbf{b}}_i^{(1)}$ and $\widehat{\mathbf{b}}_{ij}^{(2)}$ that maximize the integrands in (3.4). It can be verified that

$$\begin{aligned} &\frac{\partial g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_i, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)})}{\partial \mathbf{b}_i^{(1)}} \\ &\quad = \mathbf{Z}_i^{(1)'} \left[ \mathbf{y}_i - d'(\mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i^{(1)} \mathbf{b}_i^{(1)} + \mathbf{Z}_i^{(2)} \mathbf{b}_i^{(2)}) \right] / a(\phi) - \boldsymbol{\Psi}_1^{-1} \mathbf{b}_i^{(1)} \\ &\quad = \mathbf{Z}_i^{(1)'} \left( \mathbf{y}_i - \boldsymbol{\mu}_i \right) / a(\phi) - \boldsymbol{\Psi}_1^{-1} \mathbf{b}_i^{(1)} \end{aligned} \tag{3.5}$$

$$\begin{aligned} &\frac{\partial g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_{ij}, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)})}{\partial \mathbf{b}_{ij}^{(2)}} \\ &\quad = \mathbf{Z}_{ij}^{(2)'} \left[ \mathbf{y}_{ij} - d'(\mathbf{X}_{ij} \boldsymbol{\beta} + \mathbf{Z}_{ij}^{(1)} \mathbf{b}_i^{(1)} + \mathbf{Z}_{ij}^{(2)} \mathbf{b}_{ij}^{(2)}) \right] / a(\phi) - \boldsymbol{\Psi}_2^{-1} \mathbf{b}_{ij}^{(2)} \\ &\quad = \mathbf{Z}_{ij}^{(2)'} \left( \mathbf{y}_{ij} - \boldsymbol{\mu}_{ij} \right) / a(\phi) - \boldsymbol{\Psi}_2^{-1} \mathbf{b}_{ij}^{(2)} \end{aligned}$$

$$\begin{aligned} &\frac{\partial^2 g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_i, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)})}{\partial \mathbf{b}_i^{(1)} \partial \mathbf{b}_i^{(1)'}} \\ &\quad = -\left( \mathbf{Z}_i^{(1)'} d''\left( \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i^{(1)} \mathbf{b}_i^{(1)} + \mathbf{Z}_i^{(2)} \mathbf{b}_i^{(2)} \right) \mathbf{Z}_i^{(1)} / a(\phi) + \boldsymbol{\Psi}_1^{-1} \right) \\ &\quad = -\left( \mathbf{Z}_i^{(1)'} \frac{\mathbf{V}_i}{a^2(\phi)} \mathbf{Z}_i^{(1)} + \boldsymbol{\Psi}_1^{-1} \right), \end{aligned}$$

$$\begin{aligned} &\frac{\partial^2 g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_{ij}, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)})}{\partial \mathbf{b}_{ij}^{(2)} \partial \mathbf{b}_{ij}^{(2)'}} \\ &\quad = -\left( \mathbf{Z}_{ij}^{(2)'} d''(\mathbf{X}_{ij} \boldsymbol{\beta} + \mathbf{Z}_{ij}^{(1)} \mathbf{b}_i^{(1)} + \mathbf{Z}_{ij}^{(2)} \mathbf{b}_{ij}^{(2)}) \mathbf{Z}_{ij}^{(2)} / a(\phi) + \boldsymbol{\Psi}_2^{-1} \right) \\ &\quad = -\left( \mathbf{Z}_{ij}^{(2)'} \frac{\mathbf{V}_{ij}}{a^2(\phi)} \mathbf{Z}_{ij}^{(2)} + \boldsymbol{\Psi}_2^{-1} \right), \end{aligned}$$

$$\begin{aligned} &\frac{\partial^2 g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_{ij}, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)})}{\partial \mathbf{b}_i^{(1)} \partial \mathbf{b}_{ij}^{(2)'}} \\ &\quad = -\mathbf{Z}_{ij}^{(1)'} d''(\mathbf{X}_{ij} \boldsymbol{\beta} + \mathbf{Z}_{ij}^{(1)} \mathbf{b}_i^{(1)} + \mathbf{Z}_{ij}^{(2)} \mathbf{b}_{ij}^{(2)}) \mathbf{Z}_{ij}^{(2)} / a(\phi) \\ &\quad = -\mathbf{Z}_{ij}^{(1)'} \frac{\mathbf{V}_{ij}}{a^2(\phi)} \mathbf{Z}_{ij}^{(2)}, \end{aligned}$$

and

$$\frac{\partial^2 g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_{ij}, \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)})}{\partial \mathbf{b}_{ij}^{(2)} \partial \mathbf{b}_{ik}^{(2)\prime}} = \mathbf{0},$$

where $\mathbf{V}_i = \text{var}(\mathbf{y}_i | \mathbf{b}_i^{(1)}, \mathbf{b}_i^{(2)})$ and $\mathbf{V}_{ij} = \text{var}(\mathbf{y}_{ij} | \mathbf{b}_i^{(1)}, \mathbf{b}_{ij}^{(2)})$ represent diagonal matrices with elements $\text{var}(y_{ijk} | \mathbf{b}_i^{(1)}, \mathbf{b}_{ij}^{(2)})$. Letting $\mathbf{b}_i = (\mathbf{b}_i^{(2)\prime}, \mathbf{b}_i^{(1)\prime})'$ denote the vector of all random effects corresponding to Level 1 group $i$, it follows from the equations in (3.5) that (see the Appendix for further details)

$$\begin{aligned}
&\frac{\partial g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_i, \mathbf{b}_i)}{\partial \mathbf{b}_i} \\
&\qquad = \left[\mathbf{Z}_i^{(2)} \mathbf{Z}_i^{(1)}\right]' \left(\mathbf{y}_i - \boldsymbol{\mu}_i\right) / a(\phi) - \begin{bmatrix} \mathbf{I}_{M_i} \otimes \boldsymbol{\Psi}_2^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_1^{-1} \end{bmatrix} \mathbf{b}_i \qquad (3.6) \\
&\frac{\partial^2 g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_i, \mathbf{b}_i)}{\partial \mathbf{b}_i \partial \mathbf{b}_i'} \\
&\qquad = -\left( \left[\mathbf{Z}_i^{(2)} \mathbf{Z}_i^{(1)}\right]' \frac{\mathbf{V}_i}{a^2(\phi)} \left[\mathbf{Z}_i^{(2)} \mathbf{Z}_i^{(1)}\right] + \begin{bmatrix} \mathbf{I}_{M_i} \otimes \boldsymbol{\Psi}_2^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_1^{-1} \end{bmatrix} \right).
\end{aligned}$$

Note that the second-order partial derivatives of $g_1$ with respect to $\mathbf{b}_i$ is negative-definite and so $g_1(., \mathbf{b}_i)$ is a strictly concave function of $\mathbf{b}_i$. Therefore, there exist unique maxima $\widehat{\mathbf{b}}_i^{(1)}, \widehat{\mathbf{b}}_{ij}^{(2)}$ corresponding to the solution of $\partial\, g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_i, \mathbf{b}_i) / \partial \mathbf{b}_i |_{\mathbf{b}_i = \widehat{\mathbf{b}}_i} = \mathbf{0}$.

A Newton-Raphson algorithm for determining $\widehat{\mathbf{b}}_i$ using the expressions in (3.6) is given by

$$\begin{aligned}
\widehat{\mathbf{b}}_i^{k+1} = \widehat{\mathbf{b}}_i^k &+ \left( \left[\mathbf{Z}_i^{(2)} \mathbf{Z}_i^{(1)}\right]' \frac{\mathbf{V}_i}{a^2(\phi)} \left[\mathbf{Z}_i^{(2)} \mathbf{Z}_i^{(1)}\right] + \begin{bmatrix} \mathbf{I}_{M_i} \otimes \boldsymbol{\Psi}_2^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_1^{-1} \end{bmatrix} \right)^{-1} \\
&\times \left( \left[\mathbf{Z}_i^{(2)} \mathbf{Z}_i^{(1)}\right]' \left(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}_i^k\right) / a(\phi) - \begin{bmatrix} \mathbf{I}_{M_i} \otimes \boldsymbol{\Psi}_2^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_1^{-1} \end{bmatrix} \widehat{\mathbf{b}}_i^k \right). \qquad (3.7)
\end{aligned}$$

As in the single-level case, a more computationally efficient implementation of (3.7) is obtained by converting it into a least-squares problem. Let the augmented $\mathbf{Z}_i^{\text{aug}}$ matrix and working residuals $\mathbf{r}_i$ be defined as

$$\mathbf{Z}_i^{\text{aug}} = \begin{bmatrix} \frac{\mathbf{V}_i^{1/2}}{a(\phi)} \mathbf{Z}_i^{(2)} & \frac{\mathbf{V}_i^{1/2}}{a(\phi)} \mathbf{Z}_i^{(1)} \\ \mathbf{I}_{M_i} \otimes \boldsymbol{\Delta}_2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_1 \end{bmatrix} \quad \text{and} \quad \mathbf{r}_i = \begin{pmatrix} \mathbf{V}_i^{-1/2} \left(\mathbf{y}_i - \widehat{\boldsymbol{\mu}}_i^k\right) \\ - \begin{bmatrix} \mathbf{I}_{M_i} \otimes \boldsymbol{\Delta}_2 & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Delta}_1 \end{bmatrix} \widehat{\mathbf{b}}_i^k \end{pmatrix}, \qquad (3.8)$$

where $\boldsymbol{\Delta}_l$, satisfying $\boldsymbol{\Delta}_l' \boldsymbol{\Delta}_l = \boldsymbol{\Psi}_l^{-1}$ denotes the precision matrix at level $l$. It then follows that the increment $\widehat{\delta}_i^{k+1} = \widehat{\mathbf{b}}_i^{k+1} - \widehat{\mathbf{b}}_i^k$ corresponds to the LS solution of $\mathbf{r}_i = \mathbf{Z}_i^{\text{aug}} \boldsymbol{\delta}$.

The $\mathbf{Z}_i^{\text{aug}}$ matrix in (3.8) has a "loosely coupled" structure (Soo and Bates 1992) identical to the one encountered in multilevel linear mixed-effects (LME) models. Therefore, the

same efficient computational methods for multilevel LME models described by Pinheiro and Bates (2000) can be applied here to obtain the conditional modes $\widehat{\mathbf{b}}_i$ and the determinant of the Hessian matrix $\mathbf{Z}_i^{\text{aug}\prime}\mathbf{Z}_i^{\text{aug}}$. Following Pinheiro and Bates (2000), we initially form Q-R decompositions of the Level 2 arrays

$$\begin{bmatrix} \frac{\mathbf{V}_{ij}^{1/2}}{a(\phi)}\mathbf{Z}_{ij}^{(2)} & \frac{\mathbf{V}_{ij}^{1/2}}{a(\phi)}\mathbf{Z}_{ij}^{(1)} & \mathbf{V}_{ij}^{-1/2}(\mathbf{y}_{ij} - \widehat{\boldsymbol{\mu}}_{ij}^{k}) \\ \boldsymbol{\Delta}_2 & \mathbf{0} & -\boldsymbol{\Delta}_2\widehat{\mathbf{b}}_{ij}^{(2)k} \end{bmatrix}$$

$$= \mathbf{Q}_{ij}^{(2)} \begin{bmatrix} \mathbf{R}_{22(ij)} & \mathbf{R}_{21(ij)} & \mathbf{c}_{2(ij)} \\ \mathbf{0} & \mathbf{R}_{11(ij)} & \mathbf{c}_{1(ij)} \end{bmatrix}, \tag{3.9}$$

and then proceed to form the Level 1 Q-R decompositions

$$\begin{bmatrix} \mathbf{R}_{11(i1)} & \mathbf{c}_{1(i1)} \\ \vdots & \vdots \\ \mathbf{R}_{11(iM_i)} & \mathbf{c}_{1(iM_i)} \\ \boldsymbol{\Delta}_1 & -\boldsymbol{\Delta}_1\widehat{\mathbf{b}}_i^{(1)k} \end{bmatrix} = \mathbf{Q}_i^{(1)} \begin{bmatrix} \mathbf{R}_{11(i)} & \mathbf{c}_{1(i)} \\ \mathbf{0} & \mathbf{c}_{0(i)} \end{bmatrix}. \tag{3.10}$$

Adapting the results in Pinheiro and Bates (2000) to the the multilevel GLMM context, it can be shown that the LS problem determining the increments for $\widehat{\mathbf{b}}_i$ and $\widehat{\mathbf{b}}_{ij}$ is equivalent to minimizing

$$\left\|\mathbf{c}_{1(i)} - \mathbf{R}_{11(i)}\boldsymbol{\delta}_i^{(1)}\right\|^2 + \sum_{j=1}^{M_i}\left\|\mathbf{c}_{2(ij)} - \mathbf{R}_{22(ij)}\boldsymbol{\delta}_{ij}^{(2)} - \mathbf{R}_{21(ij)}\boldsymbol{\delta}_i^{(1)}\right\|^2,$$

which gives

$$\begin{aligned} \widehat{\mathbf{b}}_i^{(1)k+1} &= \widehat{\mathbf{b}}_i^{(1)k} + \mathbf{R}_{11(i)}^{-1}\mathbf{c}_{1(i)} \\ \widehat{\mathbf{b}}_{ij}^{(2)k+1} &= \widehat{\mathbf{b}}_{ij}^{(2)k} + \mathbf{R}_{22(ij)}^{-1}\left(\mathbf{c}_{2(ij)} - \mathbf{R}_{21(ij)}\mathbf{R}_{11(ij)}^{-1}\mathbf{c}_{1(ij)}\right) \\ \frac{1}{2}\log\left|\mathbf{Z}_i^{\text{aug}\prime}\mathbf{Z}_i^{\text{aug}}\right| &= \log\left|\mathbf{R}_{11(i)}\right| + \sum_{j=1}^{M_i}\log\left|\mathbf{R}_{22(ij)}\right|. \end{aligned} \tag{3.11}$$

Note that the matrices in (3.9) have dimensions $(n_{ij} + q_2) \times (q1 + q2 + 1)$ and the matrices in (3.10) have dimensions $(\sum_j n_{ij} + q_1) \times (q_1 + 1)$, making the calculations computationally simple and storage efficient. Also, the computations (3.11) only involve the inversion of $q_1 \times q_1$ and $q_2 \times q_2$ upper-triangular matrices, which is computationally efficient. Furthermore, the log-determinants in the right side of last equation in (3.11) are simply the sum of the logarithms of the absolute diagonal values of the corresponding upper-triangular matrices.

The Laplacian approximation to the two-level GLMM log-likelihood corresponding to the Level 1 group $i$ is then given by

$$\ell_{\text{Lap2}}(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi|\mathbf{y}_i) = \log|\boldsymbol{\Delta}_1| - \log|\mathbf{R}_{11(i)}| + M_i\log|\boldsymbol{\Delta}_2| - \sum_{j=1}^{M_i}\log|\mathbf{R}_{22(ij)}| + g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_i, \widehat{\mathbf{b}}_i^{(1)}, \widehat{\mathbf{b}}_{ij}^{(2)})$$

and the approximation for the full data is just $\ell_{\text{Lap2}}(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi|\mathbf{y}) = \sum_{i=1}^{M} \ell_{\text{Lap2}}(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi|\mathbf{y}_i)$. In the case of the binomial distribution, this simplifies to

$$\begin{aligned}\ell^{\text{Bin}}_{\text{Lap2}}(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2|\mathbf{y}_i) &= \log|\boldsymbol{\Delta}_1| - \log|\mathbf{R}_{11(i)}| - \frac{1}{2}\left\|\boldsymbol{\Delta}_1\widehat{\mathbf{b}}_i^{(1)}\right\|^2 \\ &+ M_i \log|\boldsymbol{\Delta}_2| + \sum_{j=1}^{M_i}\Big\{-\log|\mathbf{R}_{22(ij)}| + \mathbf{y}'_{ij}\left(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{(1)}\widehat{\mathbf{b}}_i^{(1)} + \mathbf{Z}_{ij}^{(2)}\widehat{\mathbf{b}}_{ij}^{(2)}\right) \\ &- \mathbf{n}'_{ij}\log\left[\mathbf{1} + \exp(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{(1)}\widehat{\mathbf{b}}_i^{(1)} + \mathbf{Z}_{ij}^{(2)}\widehat{\mathbf{b}}_{ij}^{(2)})\right] - \frac{1}{2}\left\|\boldsymbol{\Delta}_2\widehat{\mathbf{b}}_{ij}^{(2)}\right\|^2 + \sum_k \log\binom{n_{ijk}}{y_{ijk}}\Big\}\end{aligned}$$

and in the Poisson case

$$\begin{aligned}\ell^{\text{Pois}}_{\text{Lap2}}(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2|\mathbf{y}_i) &= \log|\boldsymbol{\Delta}_1| - \log|\mathbf{R}_{11(i)}| - \frac{1}{2}\left\|\boldsymbol{\Delta}_1\widehat{\mathbf{b}}_i^{(1)}\right\|^2 \\ &+ M_i \log|\boldsymbol{\Delta}_2| + \sum_{j=1}^{M_i}\Big\{ - \log|\mathbf{R}_{22(ij)}| + \mathbf{y}'_{ij}\left(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{(1)}\widehat{\mathbf{b}}_i^{(1)} + \mathbf{Z}_{ij}^{(2)}\widehat{\mathbf{b}}_{ij}^{(2)}\right) \\ &- \exp(\mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{(1)}\widehat{\mathbf{b}}_i^{(1)} + \mathbf{Z}_{ij}^{(2)}\widehat{\mathbf{b}}_{ij}^{(2)})'\mathbf{1} - \frac{1}{2}\left\|\boldsymbol{\Delta}_2\widehat{\mathbf{b}}_{ij}^{(2)}\right\|^2 - \sum_k \log(y_{ijk}!)\Big\}.\end{aligned}$$

As before, the Laplacian approximation for the full data is just the sum of the individual Laplacian approximations.

The results and methods described here for the two-level Laplacian approximation are straightforward to be extended to multilevel GLMMs with arbitrary number of levels, though the notation becomes increasingly cumbersome. The ideas of translating the estimation of $\widehat{\mathbf{b}}_i$ into an LS problem and using efficient computational methods for multilevel LME models based on Q-R decompositions of compactly stored methods carry on in the obvious way.

### 3.2 ADAPTIVE GAUSSIAN QUADRATURE APPROXIMATION

Let, as in Section 2.3, $z_i, w_i,\ i = 1, \ldots, N_{\text{GQ}}$ denote the abscissas and corresponding weights for the one-dimensional Gaussian quadrature rule. For simplicity of notation, we will assume here that the same number of abscissas is used for all levels of grouping, but it is straightforward to allow the more general case of unequal $N_{\text{GQ}}$ per level. Define $\mathbf{z}_{\mathbf{k}}^{(1)} = \left(z_{k_1}, \ldots, z_{k_{q_1}}\right)'$, and $W_{\mathbf{k}}^{(1)} = \exp(\|\mathbf{z}_{\mathbf{k}}^{(1)}\|^2)\prod_{r=1}^{q_1} w_{k_r}$ to be the Level 1 abscissas grid and its corresponding combined weights. Similarly, the Level 2 abscissas and combined weights are defined by $\mathbf{z}_{\mathbf{l}}^{(2)} = \left(z_{l_1}, \ldots, z_{l_{q_2}}\right)'$, and $W_{\mathbf{l}}^{(2)} = \exp(\|\mathbf{z}_{\mathbf{l}}^{(1)}\|^2)\prod_{s=1}^{q_2} w_{l_s}$. Using the same reasoning as in Section 2.3, combined with the results on the Q-R decompositions in (3.9) and (3.10) and defining

$$\widetilde{\mathbf{b}}_i^{(1)}(\mathbf{k}) = \widehat{\mathbf{b}}_i^{(1)} + \mathbf{R}_{11(i)}^{-1}\mathbf{z}_{\mathbf{k}}^{(1)},$$

and

$$\widetilde{\mathbf{b}}_{ij}^{(2)}(\mathbf{l}, \mathbf{k}) = \widehat{\mathbf{b}}_{ij}^{(2)} + \mathbf{R}_{22(ij)}^{-1}\left(\mathbf{z}_{\mathbf{l}}^{(2)} - \mathbf{R}_{21(ij)}\mathbf{R}_{11(ij)}^{-1}\mathbf{z}_{\mathbf{k}}^{(1)}\right),$$

it can be verified that the two-level GLMM adaptive Gaussian approximation for the $i$th Level 1 group is

$$\ell_{\mathrm{AGQ2}}\left(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi \mid \mathbf{y}_i\right) = \log|\boldsymbol{\Delta}_1| - \log\left|\mathbf{R}_{11(i)}\right| + M_i \log|\boldsymbol{\Delta}_2| - \sum_{j=1}^{M_i} \log\left|\mathbf{R}_{22(ij)}\right|$$
$$+ \log \sum_{\mathbf{k}}^{N_{\mathrm{GQ}}} \left( \exp\left[ g_{21}\left(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \phi, \mathbf{y}_i, \widetilde{\mathbf{b}}_i^{(1)}(\mathbf{k})\right)\right] W_{\mathbf{k}}^{(1)} \right.$$
$$\left. \times \prod_{j=1}^{M_i} \sum_{\mathbf{l}}^{N_{\mathrm{GQ}}} \exp\left[ g_{22}\left(\boldsymbol{\beta}, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_{ij}, \widetilde{\mathbf{b}}_i^{(1)}(\mathbf{k}), \widetilde{\mathbf{b}}_{ij}^{(2)}(\mathbf{l}, \mathbf{k})\right)\right] W_{\mathbf{l}}^{(2)} \right).$$

Note that the approximation above requires $N_{\mathrm{GQ}}^{q_1}$ evaluations of $g_{21}(.)$ and $M_i N_{\mathrm{GQ}}^{q_1+q_2}$ evaluations of $g_{22}(.)$, being computationally feasible and increasing linearly with the number of Level 2 groups. In contrast, a direct application of the AGQ approximation, without taking into account the loosely coupled structure of the LS problem discussed before, would require $N_{\mathrm{GQ}}^{q_1+M_i q_2}$ evaluations of the $g_1(.)$ function, quickly becoming computationally infeasible, even for moderate $M_i$, $q_1$, and $q_2$. As for the multilevel Laplacian approximation, extensions to an arbitrary number of levels is straightforward but notationally cumbersome.

The full data adaptive Gaussian approximation is $\ell_{\mathrm{AGQ2}}(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi|\mathbf{y}) = \sum_{i=1}^{M} \ell_{\mathrm{AGQ2}}(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi|\mathbf{y}_i)$. As in the two-level case, it is easy to show that the one point (i.e., $N_{\mathrm{GQ}} = 1$) adaptive Gaussian quadrature approximation is simply the Laplacian approximation.

In the case of the Binomial distribution, the two-level AGQ approximation simplifies to

$$\ell_{\mathrm{AGQ2}}^{\mathrm{Bin}}\left(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2 \mid \mathbf{y}_i\right) = \log|\boldsymbol{\Delta}_1| - \log\left|\mathbf{R}_{11(i)}\right| + M_i \log|\boldsymbol{\Delta}_2|$$
$$- \sum_{j=1}^{M_i} \left[ \log\left|\mathbf{R}_{22(ij)}\right| - \sum_r \log \binom{n_{ijr}}{y_{ijr}} \right]$$
$$+ \log \sum_{\mathbf{k}}^{N_{\mathrm{GQ}}} \left( \exp\left[ \mathbf{y}_i' \left\{ \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i^{(1)}\widetilde{\mathbf{b}}_i^{(1)}(\mathbf{k}) \right\} - \frac{1}{2} \left\| \boldsymbol{\Delta}_1 \widetilde{\mathbf{b}}_i^{(1)}(\mathbf{k}) \right\|^2 \right] W_{\mathbf{k}}^{(1)} \right.$$
$$\times \prod_{j=1}^{M_i} \sum_{\mathbf{l}}^{N_{\mathrm{GQ}}} \exp\left[ \mathbf{y}_{ij}' \mathbf{Z}_{ij}^{(2)} \widetilde{\mathbf{b}}_{ij}^{(2)}(\mathbf{l}, \mathbf{k}) - \mathbf{n}_{ij}' \right.$$
$$\log\left( \mathbf{1} + \exp\left\{ \mathbf{X}_{ij}\boldsymbol{\beta} + \mathbf{Z}_{ij}^{(1)}\widetilde{\mathbf{b}}_i^{(1)}(\mathbf{k}) + \mathbf{Z}_{ij}^{(2)}\widetilde{\mathbf{b}}_{ij}^{(2)}(\mathbf{l}, \mathbf{k}) \right\} \right)$$
$$\left. \left. - \frac{1}{2} \left\| \Delta_2 \widetilde{\mathbf{b}}_{ij}^{(2)}(\mathbf{l}, \mathbf{k}) \right\|^2 \right] W_{\mathbf{l}}^{(2)} \right)$$

and in the Poisson case

$$\ell_{\mathrm{AGQ2}}^{\mathrm{Pois}}\left(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2 \mid \mathbf{y}_i\right) = \log|\boldsymbol{\Delta}_1| - \log\left|\mathbf{R}_{11(i)}\right| + M_i \log|\boldsymbol{\Delta}_2|$$
$$- \sum_{j=1}^{M_i} \left[ \log\left|\mathbf{R}_{22(ij)}\right| + \sum_r \log(y_{ijr}!) \right]$$

$$+\log\sum_{\mathbf{k}}^{N_{GQ}}\left(\exp\left[\mathbf{y}_i'\left\{\mathbf{X}_i\boldsymbol{\beta}+\mathbf{Z}_i^{(1)}\widetilde{\mathbf{b}}_i^{(1)}(\mathbf{k})\right\}-\frac{1}{2}\left\|\Delta_1\widetilde{\mathbf{b}}_i^{(1)}(\mathbf{k})\right\|^2\right]W_{\mathbf{k}}^{(1)}\right.$$
$$\times\prod_{j=1}^{M_i}\sum_{\mathbf{l}}^{N_{GQ}}\exp\left\{\mathbf{y}_{ij}'\mathbf{Z}_{ij}^{(2)}\widetilde{\mathbf{b}}_{ij}^{(2)}(\mathbf{l},\mathbf{k})-\exp\left[\mathbf{X}_{ij}\boldsymbol{\beta}+\mathbf{Z}_{ij}^{(1)}\widetilde{\mathbf{b}}_i^{(1)}(\mathbf{k})+\mathbf{Z}_{ij}^{(2)}\widetilde{\mathbf{b}}_{ij}^{(2)}(\mathbf{l},\mathbf{k})\right]'\mathbf{1}\right.$$
$$\left.\left.-\frac{1}{2}\left\|\Delta_2\widetilde{\mathbf{b}}_{ij}^{(2)}(\mathbf{l},\mathbf{k})\right\|^2\right\}W_{\mathbf{l}}^{(2)}\right).$$

## 4. EVALUATING THE APPROXIMATIONS VIA SIMULATIONS

This section illustrates the use and evaluate the accuracy of the Laplacian and AGQ approximations using simulated data from binomial and Poisson GLMMs. To evaluate the impact of the number of quadrature points on the AGQ approximation, rules with 3, 5, and 7 points denoted, respectively, AGQ3, AGQ5, and AGQ7, are considered in the simulations.

The first example we consider is given by a set of 100 simulated datasets produced by Rodriguez and Goldman (1995) and which has been used by them to investigate the statistical properties of different GLMM approximation algorithms. The simulated data follow the same hierarchical structure as the data from a 1987 National Survey of Maternal and Child Health in Guatemala. This was a multistage survey on women aged 15–44 years living in clusters of communities. The goal was to investigate the determinants of use of modern prenatal care versus conventional care during pregnancy. A total of 2,449 births occurred to mothers who reported having obtained some type of prenatal care. These births were associated with 1,558 families in 161 communities. The sample sizes per community range from 1 to 50 with a mean of 15 children. The response variable is a binary indicator $y_{ijk}$ taking the value 1 when modern prenatal care was used with the $k$th child within the $j$th family within the $i$th community and 0 when conventional care was used. The GLMM model used to produce the simulations was

$$\begin{aligned}\eta_{ijk} &= \log\left(\frac{\mu_{ijk}}{1-\mu_{ijk}}\right)=\beta_0+\beta_1x_{1,i}+\beta_2x_{2,ij}+\beta_3x_{3,ijk}+b_i^{(1)}+b_{ij}^{(2)}\\ \mu_{ijk} &= E\left(y_{ijk}|b_i^{(1)},b_{ij}^{(2)}\right)\quad b_i^{(1)}\overset{\text{ind}}{\sim}\mathcal{N}\left(0,\sigma_1^2\right)\quad b_{ij}^{(2)}\overset{\text{ind}}{\sim}\mathcal{N}\left(0,\sigma_2^2\right).\end{aligned}\tag{4.1}$$

The community, family, and child covariates, $x_{1,i}$, $x_{2,ij}$, and $x_{3,ijk}$, respectively, were generated as part of the simulation. The community random effects $b_i^{(1)}$ are independent of the family random effects $b_{ij}^{(2)}$. The parameter values used in the simulations were, respectively, $\beta_0=0.665, \beta_1=\beta_2=\beta_3=\sigma_1^2=\sigma_2^2=1$. The values chosen for the variance components $\sigma_1^2$ and $\sigma_2^2$ result in fairly high variability in the responses.

Several different estimation approaches have been investigated and compared in the literature to analyze the Rodriguez and Goldman simulated datasets. For comparison with the Laplacian and AGQ methods, we include here two of the most popular of such methods: PQL and restricted PQL (REPQL) (Breslow and Clayton 1993). The PQL and REPQL implementations used here allow for a dispersion parameter that is estimated with the rest

Table 1. Average Values of Fixed Effects and Standard Deviation Estimates for the Rodriguez and Goldman Simulated Binomial Data Under Different Estimation Methods

| | *Intercept* | *Community* | *Family* | *Child* | *Community* | *Family* |
|---|---|---|---|---|---|---|
| | $\beta_0$ | $\beta_1$ | $\beta_2$ | $\beta_3$ | $\sigma_1$ | $\sigma_2$ |
| *Method* | *0.665* | *1.000* | *1.000* | *1.000* | *1.000* | *1.000* |
| RG | NA | 0.771 | 0.744 | 0.738 | 0.732 | 0.103 |
| PQL | 0.632 | 0.987 | 0.942 | 0.975 | 0.921 | 1.423 |
| REPQL | 0.633 | 0.989 | 0.944 | 0.977 | 0.934 | 1.429 |
| Laplacian | 0.621 | 0.951 | 0.909 | 0.903 | 0.881 | 0.583 |
| AGQ3 | 0.661 | 1.014 | 0.969 | 0.964 | 0.944 | 0.898 |
| AGQ5 | 0.676 | 1.037 | 0.989 | 0.982 | 0.972 | 0.975 |
| AGQ7 | 0.677 | 1.039 | 0.990 | 0.984 | 0.974 | 0.979 |

of the parameters in the model, giving them more flexibility in the estimation of the variance components. By contrast, the Laplacian and AGQ approximations are true likelihood methods and so do not include the estimation of dispersion parameters (which are assumed to be equal to 1 in the binomial and Poisson cases). Table 1 gives the averages of the fixed effects and standard deviation estimates for the different methods over the 100 simulated datasets. The original results from Rodriguez and Goldman (1995) (RG) are also included for comparison.

The results in Table 1 indicate that RG, PQL, REPQL, and Laplacian all lead to severely biased standard deviation estimates. The fixed effects estimates are also biased under these methods, being all underestimated, but relatively less biased than the estimated standard deviations. The AGQ quadrature approximations have considerably smaller bias, with their performance improving with the number of quadrature points, as expected. There is little difference between the AGQ5 and AGQ7 approximations, suggesting that five quadrature points provides sufficient accuracy for this example. Computations of 95% confidence intervals (assuming asymptotic normality of the maximum likelihood estimates) for the parameter estimates in the simulated datasets, not included here, resulted in coverages close to the nominal level for all parameters using AGQ5 and AGQ7, but below 90% for the standard deviations using the other methods, and even below 10% for $\sigma_2$, when using PQL, REPQL and Laplacian.

Figure 1 shows the estimated densities of the fixed effects and standard deviation estimates corresponding to the REPQL, Laplacian, AGQ3, and AGQ5 methods. The density plots indicate that the estimation of the standard deviations is what discriminates the performance of the methods, in particular the estimation of $\sigma_2$. This is consistent with the observed coverage of the asymptotic confidence intervals for the variance components, mentioned above. It is also clear from Figure 1 that the superiority of AGQ5 over the other methods is due to a reduction in bias and not a reduction in variability in the estimates.

The second simulation example considered here focuses on the Laplacian and AGQ approximations, using a two-level binomial GLMM with exactly the same model structure as (4.1) to investigate the impact of the sample sizes at different levels of nesting on the precision of the estimates. Let $L_3|L_2|L_1$ represent the multilevel data structure with $L_1$ units at the first level, a total of $L_2$ units at the second level nested within the first level units, and a total of $L_3$ observations collected within the second level units. Two such data

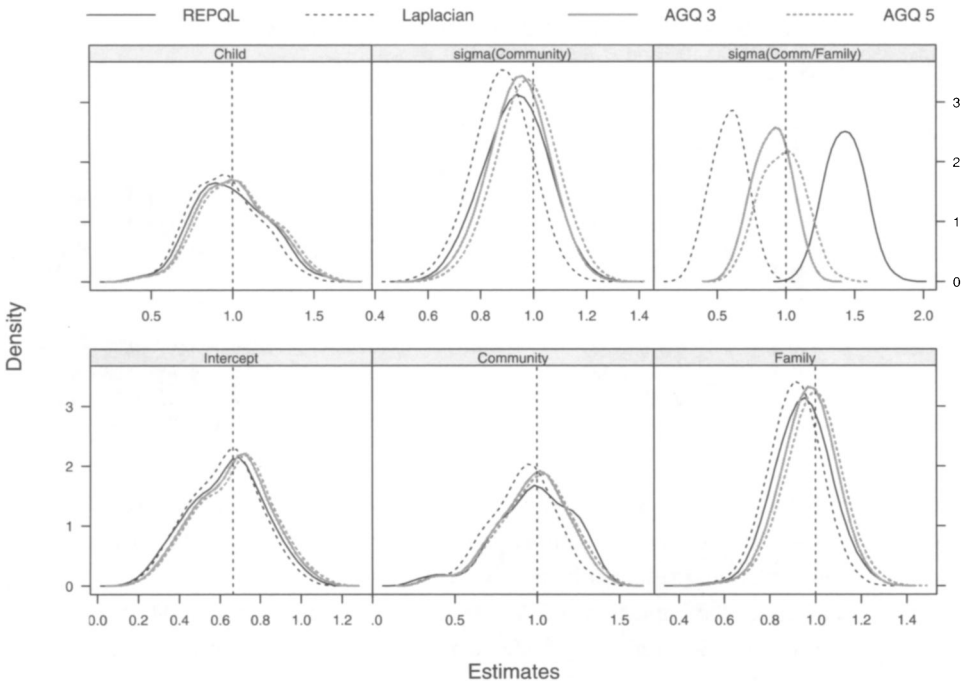


*Figure 1. Estimated densities of fixed effects and standard deviation estimates corresponding to REPQL, Laplacian, AGQ3, and AGQ5. Dashed vertical lines indicate the true parameter values used to simulate the data.*

configurations are considered here: $A = 900|300|100$ and $B = 1800|450|150$, with the second level units being randomly allocated to the first level units, to simulate imbalance, but with a fixed number of replicates being assigned to each of the second levels units (3 and 4, respectively, for the $A$ and $B$ configurations). A total of 200 simulated datasets with configurations $A$ and $B$ are used to investigate the performance of the approximations. The covariates in the multilevel logistic GLMM are independently simulated as follows: $x_{1i}$ is a binary variable taking values 0 or 1 that splits the first level units into two groups of equal size, $x_{2ij} \overset{\text{ind}}{\sim} N(0,1)$, and $x_{3ijk} \overset{\text{ind}}{\sim} U(-1,1)$. The parameter values used in the simulations are $\beta_0 = -0.7$, $\beta_1 = -1$, $\beta_2 = 1$, $\beta_3 = -0.5$, and $\sigma_1 = \sigma_2 = 1$. Table 2 summarizes the simulation results, listing the means and standard deviations of the parameter estimates obtained with the different approximations.

All methods produce relatively little or no bias in the estimation of the fixed effects, under the model and data configurations assumed in the simulations. There appears to be some downward bias in the estimation of the variance components, with the magnitude of the bias decreasing with the number of quadrature points. The ACG5 and ACG7 approximations give nearly unbiased estimates, with very similar performances. Increasing the sample size from configuration $A$ to configuration $B$ leads to an overall reduction in the magnitude of the fixed effects bias, but less so with the variance components, especially for the Laplacian and the AGQ3 approximations. The standard deviations of the estimates are about the same across all methods, for the same data configuration. The increase in sample

Table 2. Monte Carlo Means and Standard Deviations (in parentheses) of the Parameter Estimates in the Multilevel GLMM Logistic Model, Under Different Approximations and for Data Configurations $A = 900|300|100$ and $B = 1800|450|150$.

| Method | $\beta_0$ $-0.700$ | $\beta_1$ $-1.000$ | $\beta_2$ $1.000$ | $\beta_3$ $-0.500$ | $\sigma_1$ $1.000$ | $\sigma_2$ $1.000$ |
|---|---|---|---|---|---|---|
| | | | *Configuration A* | | | |
| Lapl. | −0.670 (0.231) | −1.055 (0.327) | 0.996 (0.141) | −0.488 (0.177) | 0.941 (0.186) | 0.864 (0.217) |
| AGQ3 | −0.673 (0.231) | −1.061 (0.328) | 1.002 (0.138) | −0.493 (0.178) | 0.958 (0.182) | 0.941 (0.189) |
| AGQ5 | −0.683 (0.235) | −1.076 (0.334) | 1.016 (0.142) | −0.499 (0.181) | 0.979 (0.190) | 0.997 (0.211) |
| AGQ7 | −0.684 (0.235) | −1.077 (0.334) | 1.017 (0.142) | −0.499 (0.181) | 0.980 (0.191) | 1.000 (0.213) |
| | | | *Configuration B* | | | |
| Lapl. | −0.702 (0.173) | −0.975 (0.239) | 0.980 (0.105) | −0.487 (0.113) | 0.960 (0.140) | 0.894 (0.138) |
| AGQ3 | −0.701 (0.172) | −0.973 (0.238) | 0.978 (0.103) | −0.488 (0.113) | 0.963 (0.138) | 0.930 (0.125) |
| AGQ5 | −0.710 (0.175) | −0.985 (0.241) | 0.989 (0.105) | −0.492 (0.114) | 0.985 (0.143) | 0.981 (0.137) |
| AGQ7 | −0.711 (0.175) | −0.986 (0.241) | 0.990 (0.105) | −0.492 (0.115) | 0.988 (0.143) | 0.985 (0.139) |

size from configuration $A$ to configuration $B$ leads to substantial gains in efficiency, more noticeably among the fixed effects estimates. The observed coverage of the estimated 95% confidence intervals, not presented here, is very close to the nominal value. The only exceptions occurred for the Laplacian and AGQ3 approximations with regard to the variance components: 93% coverage for $\sigma_1$ and 91% coverage for $\sigma_2$. These simulation results reinforce that previous indication that a fit with five quadrature points gives sufficient accuracy for the AGQ approximation.

The simulation studies discussed above illustrate cases in which (RE)PQL and the Laplacian approximation lead to biased estimates of variance components and, to a lesser degree, also of fixed effects. In these cases, at least five quadrature points were required for the AGQ approximation to produce nearly unbiased estimates. The poor bias performance of the (RE)PQL and Laplacian methods in these cases is associated with the data-model configuration (binary responses with few observations per second-level unit), which leads to severe inaccuracies in the underlying approximations used for each method. There exist, of course, data-model configurations for GLMMs in which (RE)PQL and the Laplacian approximation produce nearly unbiased estimates for variance components and fixed effects. To illustrate one of such cases, we consider a third and last simulation example in which the response $y_{ij}$ is described by a single-level Poisson GLMM with logarithm link function and two covariates.

$$\begin{aligned} \eta_{ij} &= \log(\mu_{ij}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2ij} + b_i \\ \mu_{ij} &= E(y_{ij}|b_i),\ b_i \sim N(0, \sigma^2). \end{aligned}$$

The covariates are independently simulated as follows: $x_{1i}$ is a binary variable taking values 0 or 1 that splits the first level units into two groups of equal size and $x_{2ij} \overset{\text{ind}}{\sim} N(0, 1)$. The parameter values used in the simulations are $\beta_0 = 3$, $\beta_1 = -0.3$, $\beta_2 = 0.3$, and $\sigma = 1$. A total of 200 simulated datasets with data configuration 600|300 (with two replicates per first level unit) are used. Table 3 summarizes the simulation results, listing the means and standard deviations of the parameter estimates obtained with the different approximations. In this case, all estimation methods considered lead to nearly unbiased estimates for the

Table 3. Monte Carlo Means and Standard Deviations (in parentheses) of the Parameter Estimates in the Multilevel GLMM Poisson Model, under Different Approximations and for Data Configuration 600|300

| *Method* | $\beta_0$ *3.000* | $\beta_1$ *−0.300* | $\beta_2$ *0.300* | $\sigma_1$ *0.300* |
|---|---|---|---|---|
| REPQL | 3.012 (0.028) | −0.300 (0.038) | 0.299 (0.012) | 0.295 (0.018) |
| Lapl. | 3.001 (0.029) | −0.302 (0.039) | 0.299 (0.012) | 0.302 (0.072) |
| AGQ3 | 3.001 (0.029) | −0.302 (0.039) | 0.299 (0.012) | 0.302 (0.072) |
| AGQ5 | 3.001 (0.029) | −0.302 (0.039) | 0.299 (0.012) | 0.303 (0.072) |
| AGQ7 | 3.001 (0.029) | −0.302 (0.039) | 0.299 (0.012) | 0.303 (0.072) |

fixed effects and the variance component, with little being gained by using the more accurate AGQ approximations.

An important consideration when using the different GLMM estimation methods, as the ones described in this article, is the relative computational performance associated with them. Of course, measurements of computational performance are a reflection of implementations of the algorithms, and not just the algorithms themselves. They are also highly dependent on the platform and software used to run the programs. For the purpose of illustration of the relative performances of the REPQL, Laplacian, and AGQ quadrature approximations in GLMM estimation, we report, in Table 4 the average CPU times (in minutes) per simulated dataset to fit the two-level binomial GLMM of Rodriguez and Goldman (1995) and the single-level Poisson GLMM.

These CPU times refer to an S-Plus implementation of all algorithms (using calls to C++ compiled code) which was run under a fairly outdated Sun Unix Ultra-2 with dual 168 MHz processors, 256 Mb of RAM, and 825 Mb of swap space. Because REPQL does not require integration over the random effects, not surprisingly it is the fastest algorithm. For AGQ estimation, recalling that the Laplacian approximation is just a one-point AGQ approximation, the speed is approximately linear in the number of quadrature points. The AGQ5 approximation, which gives nearly unbiased estimates in the binomial example, uses about 12 times more CPU time than REPQL. For the single-level Poisson GLMM example, the linear trend is quite flat.

For a concrete comparison between AGQ methods, recall that an implementation ignoring the loosely coupled structure of the underlying LS problem would require, in the case of a two-level model, $N_{\mathrm{GQ}}^{q_1+M_i q_2}$ function evaluations for first-level unit $i$. Note that this increases exponentially with the number of second-level units, $M_i$. The AGQ implementation proposed here, by contrast, requires $N_{\mathrm{GQ}}^{q_1}(1 + M_i N_{\mathrm{GQ}}^{q_2})$ function evaluations for first-level unit $i$, which increases linearly with $M_i$. In the RG example, with an average

Table 4. Average CPU Times (in minutes) per Simulated Dataset to Fit the Two-Level Binomial GLMM in the Rodriguez and Goldman Example and the Single-Level Poisson GLMM

| | *REPQL* | *Laplacian* | *AGQ3* | *AGQ5* | *AGQ7* |
|---|---|---|---|---|---|
| Binomial (RG) | 1.474 | 7.667 | 13.317 | 17.780 | 23.686 |
| Poisson | 0.188 | 1.259 | 1.462 | 1.462 | 1.549 |

of about 10 families per community, a five-point AGQ approximation requires around 255 function evaluations per community using our proposed implementation versus $5^{11} \simeq 48.83$ million function evaluations using a single-level AGQ implementation.

More generally, the processing time to obtain the GLMM estimates will be influenced by the complexity of the model being fitted, the size and noisiness of the data, the implementation of the algorithms, and, of course, hardware characteristics (e.g., processor speed, available RAM). Halving of the CPU time for the AGQ approximations were observed in experiments conducted on a faster Xeon-based machine with 2 GHz and 2Gb RAM running the MS Windows Server 2003.

## 5. CONCLUSIONS AND FURTHER THOUGHTS

This article describes efficient Laplacian and adaptive Gaussian quadrature algorithms for likelihood estimation in the context of generalized linear mixed-effects models. The proposed algorithms extend to GLMMs the methods developed by Pinheiro and Bates (2000) for highly efficient likelihood estimation in multilevel linear mixed-effects models. The resulting Laplacian and AGQ algorithms inherit the computational and memory-usage efficiency of the LME algorithms of Pinheiro and Bates, and can be implemented using simple modifications of existing software for LME estimation, covering both single- and multilevel GLMMs. By increasing the number of quadrature points, the AGQ approximation can be made as accurate as desired (and computationally feasible). Simulation examples are included to illustrate the accuracy and use of the Laplacian and AGQ approximations, comparing them to commonly used estimation methods for GLMMs.

The proposed algorithms are designed to scale up efficiently to multilevel GLMMs with an arbitrary number of *nested* random effects. As mixed-effects models become increasingly used in practice, applications with more complex random effects structures, such as crossed random effects and combinations of nested/crossed random effects, are becoming more frequent. Extensions of the methods described here are needed to produce efficient algorithms for such cases. Because of the non-nested structure of the underlying regression matrices in these cases, efficient sparse-matrix algorithms will likely be needed.

As mentioned, the Laplacian and AGQ algorithms here described are designed for true likelihood estimation, which does not include dispersion parameters for the binomial and Poisson distributions. A useful practical extension of these algorithms would be to allow for over-dispersed binomial and Poisson GLMMs, by extending the objective function to be maximized and appropriately adapting the underlying computational methods in the approximations.

## APPENDIX

We describe here in more detail the derivations leading to Equation (3.6) in Section 3.1. From the definition of $\partial g_1 / \partial \mathbf{b}_i$ and the equations in (3.5), the vector of partial derivatives

can be expressed as

$$\frac{\partial g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_i, \mathbf{b}_i)}{\partial \mathbf{b}_i} = \begin{bmatrix} \frac{\partial g_1}{\partial \mathbf{b}_{i1}^{(2)}} \\ \vdots \\ \frac{\partial g_1}{\partial \mathbf{b}_{iM_i}^{(2)}} \\ \frac{\partial g_1}{\partial \mathbf{b}_i^{(1)}} \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{Z}_{i1}^{(2)\prime}\left(\mathbf{y}_{i1} - \boldsymbol{\mu}_{i1}\right)/a(\phi) - \boldsymbol{\Psi}_2^{-1}\mathbf{b}_{i1}^{(2)} \\ \vdots \\ \mathbf{Z}_{iM_i}^{(2)\prime}\left(\mathbf{y}_{iM_i} - \boldsymbol{\mu}_{iM_i}\right)/a(\phi) - \boldsymbol{\Psi}_2^{-1}\mathbf{b}_{iM_i}^{(2)} \\ \mathbf{Z}_i^{(1)\prime}\left(\mathbf{y}_i - \boldsymbol{\mu}_i\right)/a(\phi) - \boldsymbol{\Psi}_1^{-1}\mathbf{b}_i^{(1)} \end{bmatrix}.$$

Using simple linear algebra results, the vector on the right of the equation above can be re-expressed as

$$\begin{bmatrix} \mathbf{Z}_{i1}^{(2)\prime} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{Z}_{iM_i}^{(2)\prime} \\ & \mathbf{Z}_i^{(1)\prime} & \end{bmatrix} \frac{\mathbf{y}_i - \boldsymbol{\mu}_i}{a(\phi)} - \begin{bmatrix} \boldsymbol{\Psi}_2^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \boldsymbol{\Psi}_2^{-1} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\Psi}_1^{-1} \end{bmatrix} \mathbf{b}_i$$

$$= \left[\mathbf{Z}_i^{(2)}\mathbf{Z}_i^{(1)}\right]'\left(\mathbf{y}_i - \boldsymbol{\mu}_i\right)/a(\phi) - \begin{bmatrix} \mathbf{I}_{M_i} \otimes \boldsymbol{\Psi}_2^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_1^{-1} \end{bmatrix} \mathbf{b}_i,$$

as presented on the right hand side of the first equation in (3.6).

Similarly, using the last set of equations in (3.5), the matrix of second-order partial derivatives $\partial^2 g_1/\partial \mathbf{b}_i \partial \mathbf{b}_i'$ can be expressed as

$$\frac{\partial^2 g_1(\boldsymbol{\beta}, \boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \phi, \mathbf{y}_i, \mathbf{b}_i)}{\partial \mathbf{b}_i \partial \mathbf{b}_i'} = \begin{bmatrix} \frac{\partial^2 g_1}{\partial \mathbf{b}_{i1}^{(2)} \partial \mathbf{b}_{i1}^{(2)\prime}} & \mathbf{0} & \cdots & \frac{\partial^2 g_1}{\partial \mathbf{b}_{i1}^{(2)} \partial \mathbf{b}_i^{(1)\prime}} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \frac{\partial^2 g_1}{\partial \mathbf{b}_{iM_i}^{(2)} \partial \mathbf{b}_{iM_i}^{(2)\prime}} & \frac{\partial^2 g_1}{\partial \mathbf{b}_{iM_i}^{(2)} \partial \mathbf{b}_i^{(1)\prime}} \\ \frac{\partial^2 g_1}{\partial \mathbf{b}_i^{(1)} \partial \mathbf{b}_{i1}^{(2)\prime}} & \cdots & \frac{\partial^2 g_1}{\partial \mathbf{b}_i^{(1)} \partial \mathbf{b}_{iM_i}^{(2)\prime}} & \frac{\partial^2 g_1}{\partial \mathbf{b}_i^{(1)} \partial \mathbf{b}_i^{(1)\prime}} \end{bmatrix}$$

$$= -\begin{bmatrix} \mathbf{Z}_{i1}^{(2)\prime}\frac{\mathbf{V}_{i1}}{a^2(\phi)}\mathbf{Z}_{i1}^{(2)} + \boldsymbol{\Psi}_2^{-1} & \mathbf{0} & \cdots & \mathbf{Z}_{i1}^{(2)\prime}\frac{\mathbf{V}_{i1}}{a^2(\phi)}\mathbf{Z}_{i1}^{(1)} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \cdots & \mathbf{Z}_{iM_i}^{(2)\prime}\frac{\mathbf{V}_{iM_i}}{a^2(\phi)}\mathbf{Z}_{iM_i}^{(2)} + \boldsymbol{\Psi}_2^{-1} & \mathbf{Z}_{iM_i}^{(2)\prime}\frac{\mathbf{V}_{iM_i}}{a^2(\phi)}\mathbf{Z}_{iM_i}^{(1)} \\ \mathbf{Z}_{i1}^{(1)\prime}\frac{\mathbf{V}_{i1}}{a^2(\phi)}\mathbf{Z}_{i1}^{(2)} & \cdots & \mathbf{Z}_{iM_i}^{(1)\prime}\frac{\mathbf{V}_{iM_i}}{a^2(\phi)}\mathbf{Z}_{iM_i}^{(2)} & \mathbf{Z}_i^{(1)\prime}\frac{\mathbf{V}_i}{a^2(\phi)}\mathbf{Z}_i^{(1)} + \boldsymbol{\Psi}_1^{-1} \end{bmatrix}.$$

It then follows from simple properties of matrix algebra that the expression above for the second-order partial derivatives matrix can be simplified to

$$\partial^2 g_1/\partial \mathbf{b}_i \partial \mathbf{b}_i' = -\left( \left[\mathbf{Z}_i^{(2)}\mathbf{Z}_i^{(1)}\right]' \frac{\mathbf{V}_i}{a^2(\phi)} \left[\mathbf{Z}_i^{(2)}\mathbf{Z}_i^{(1)}\right] + \begin{bmatrix} \mathbf{I}_{M_i} \otimes \boldsymbol{\Psi}_2^{-1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Psi}_1^{-1} \end{bmatrix} \right),$$

as presented in the second equation of (3.6).

## ACKNOWLEDGMENTS

The authors thank Richard Gates, who developed the programs implementing the GLMM Laplacian and AGQ approximations in S-Plus and contributed with many useful ideas and enhancements to the algorithms. We would also like to thank Dr. German Rodriguez for providing us with the simulated data from Rodriguez and Goldman (1995). This research was partially supported by grants from U.S. Army Research Office DAMD17-02-C-0114 and NIH CA83512 and CA88754.

*[Received March 2004. Revised March 2005.]*

## REFERENCES

Abramowitz, M., and Stegun, I. A. (1964), *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, New York: Dover.

Breslow, N. E., and Clayton, D. G. (1993), "Approximate Inference in Generalized Linear Mixed Models," *Journal of the American Statistical Association,* 88, 9–25.

Clarkson, D. B., and Zhan, Y. (2002), "Using Spherical–Radial Quadrature to Fit Generalized Linear Mixed Effects Models," *Journal of Computational and Graphical Statistics*, 11, 639–659.

Davis, P. J., and Rabinowitz, P. (1984), *Methods of Numerical Integration*, New York: Academic Press.

Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–1339.

Golub, G. H. (1973), "Some Modified Matrix Eigenvalue Problems," *SIAM Review*, 15, 318–334.

Gloub, G. H., and Welsch, J. H. (1969), "Calculation of Gaussian Quadrature Rules," *Mathematical Computing*, 23, 221–230.

Hox, J. (2002), *Multilevel Analysis*, Mahwah, NJ: Lawrence Erlbaum Associates.

Laird, N. M., and Ware, J. H. (1982), "Random-Effects Models for Longitudinal Data," *Biometrics*, 38, 963–974.

Leonard, T., Hsu, J. S. J., and Tsui, K. W. (1989), "Bayesian Marginal Inference," *Journal of the American Statistical Association*, 84, 1051–1058.

Leyland, A. H., and Goldstein, H. (eds.) (2001), *Multilevel Modelling of Health Statistics*, New York: Wiley.

Lindstrom, M. J., and Bates, D. M. (1988), "Newton–Raphson and EM Algorithms for Linear Mixed-Effects Models for Repeated-Measurements Data," *Journal of the American Statistical Association*, 83, 1014–1022.

Lindstrom, M. J., and Bates, D. M. (1990), "Nonlinear Mixed-Effects Models for Repeated-Measures Data," *Biometrics,* 46, 673–687.

Littell, R. C., Milliken, G. A., Stroup, W. W., and Wolfinger, R. D. (1996), *SAS System for Mixed Models*, Cary, NC: SAS Institute, Inc.

McCulloch and Searle (2001), *Generalized, Linear, and Mixed Models*, New York: Wiley.

Pinheiro, J. C., and Bates, D. M. (1995), "Approximations to the Log-likelihood Function in the Nonlinear Mixed Effects Model," *Journal of Computational and Graphical Statistics*, 4, 12–35.

——— (2000), *Mixed-Effects Models in S and S-Plus*, New York: Springer-Verlag.

Raudenbush, S. W., and Bryk, A. S. (2002), *Hierarchical Linear Models* (2nd ed.), Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., Yang, M. Y., and Yosef, M. (2000), "Maximum Likelihood for Generalized Linear Models with Nested Random Effects via High-order, Multivariate Laplace Approximations," *Journal of Computational and Graphical Statistics*, 9, 141–157.

Rodriguez, G., and Goldman, N. (1995), "An Assessment of Estimation Procedures for Multilevel Models with Binary Responses," *Journal of the Royal Statistical Society*, Series A, 158, 73–90.

Soo, Y.-W., and Bates, D. M. (1992), "Loosely Coupled Nonlinear Least Squares," *Computational Statistics and Data Analysis*, 14, 249–259.

Tierney, L., and Kadane, J. B. (1986), "Accurate Approximation for Posterior Moments and Marginal Densities," *Journal of the American Statistical Association*, 81, 82–86.

Thisted, R. A. (1988), *Elements of Statistical Computing*, London: Chapman & Hall.